

2530
1026
#11/2
6420
TECH CENTER 1600/2900
DEC 1 8 2001
RECEIVED

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/973,025

DATE: 11/30/2001
TIME: 16:06:22

Input Set : N:\Crf3\RULE60\09973025.txt
Output Set: N:\CRF3\11302001\I973025.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: MAERTENS, GEERT

7 BOSMAN, FONS

8 DE MARTYNOFF, GUY

9 BUYSE, MARIE-ANGE

11 (ii) TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

12 PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

14 (iii) NUMBER OF SEQUENCES: 111

16 (iv) CORRESPONDENCE ADDRESS:

17 (A) ADDRESSEE: NIXON & VANDERHYE P.C.

18 (B) STREET: 1100 NORTH GLEBE ROAD

19 (C) CITY: ARLINGTON

20 (D) STATE: VIRGINIA

21 (E) COUNTRY: U.S.A.

22 (F) ZIP: 22201-4714

24 (v) COMPUTER READABLE FORM:

25 (A) MEDIUM TYPE: Floppy disk

26 (B) COMPUTER: IBM PC compatible

27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

30 (vi) CURRENT APPLICATION DATA:

C--> 31 (A) APPLICATION NUMBER: US/09/973,025

C--> 32 (B) FILING DATE: 10-Oct-2001

33 (C) CLASSIFICATION:

36 (vii) PRIOR APPLICATION DATA:

37 (A) APPLICATION NUMBER: US 08/612,973

38 (B) FILING DATE: 11-MAR-1996

41 (viii) ATTORNEY/AGENT INFORMATION:

42 (A) NAME: BYRNE, THOMAS E.

43 (B) REGISTRATION NUMBER: 32,205

44 (C) REFERENCE/DOCKET NUMBER: 1487-10

46 (ix) TELECOMMUNICATION INFORMATION:

47 (A) TELEPHONE: (703) 816-4000

48 (B) TELEFAX: (703) 816-4100

52 (2) INFORMATION FOR SEQ ID NO: 1:

54 (i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 21 base pairs

56 (B) TYPE: nucleic acid

57 (C) STRANDEDNESS: single

58 (D) TOPOLOGY: linear

60 (ii) MOLECULE TYPE: cDNA

62 (iii) HYPOTHETICAL: NO

C--> 64 (iv) ANTI-SENSE: NO

69 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

71 GGCATGCAAG CTTAATTAAT T

73 (2) INFORMATION FOR SEQ ID NO: 2:

ENTERED

21

RECEIVED
 DEC 18 2001
 TECH CENTER 1600/2900

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/973,025

DATE: 11/30/2001

TIME: 16:06:22

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

```

75      (i) SEQUENCE CHARACTERISTICS:
76          (A) LENGTH: 68 base pairs
77          (B) TYPE: nucleic acid
78          (C) STRANDEDNESS: single
79          (D) TOPOLOGY: linear
81      (ii) MOLECULE TYPE: cDNA
83      (iii) HYPOTHETICAL: NO
C--> 85      (iv) ANTI-SENSE: NO
89      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
91 CCGGGGAGGC CTGCACGTGA TCGAGGGCAG ACACCATCAC CACCATCACT AATAGTTAAT      60
93 TAACTGCA                                                                    68
95 (2) INFORMATION FOR SEQ ID NO: 3:
97      (i) SEQUENCE CHARACTERISTICS:
98          (A) LENGTH: 642 base pairs
99          (B) TYPE: nucleic acid
100         (C) STRANDEDNESS: single
101         (D) TOPOLOGY: linear
103      (ii) MOLECULE TYPE: cDNA
105      (iii) HYPOTHETICAL: NO
C--> 107      (iv) ANTI-SENSE: NO
110      (ix) FEATURE:
111          (A) NAME/KEY: CDS
112          (B) LOCATION: 1..639
114      (ix) FEATURE:
115          (A) NAME/KEY: mat_peptide
116          (B) LOCATION: 1..636
119      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
121 ATG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTA CTG TCC TGT      48
122 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
123 1      5      10      15
125 CTG ACC ATT CCA GCT TCC GCT TAT GAG GTG CGC AAC GTG TCC GGG ATG      96
126 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
127      20      25      30
129 TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA      144
130 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
131      35      40      45
133 GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT CGG GAG      192
134 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
135      50      55      60
137 AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG CTC GCA GCT      240
138 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
139 65      70      75      80
141 AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC GTC GAT TTG      288
142 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
143      85      90      95
145 CTC GTT GGG GCG GCT GCT CTC TGT TCC GCT ATG TAC GTG GGG GAT CTC      336
146 Leu Val Gly Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
147      100      105      110
149 TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC TCG CCT CGC      384

```

RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/973,025

TIME: 16:06:22

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

```

150 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
151      115      120      125
153 CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT CCC GGC CAC      432
154 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
155      130      135      140
157 ATA ACA GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC TGG TCG CCT      480
158 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
159 145      150      155      160
161 ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA CAA GCT GTC      528
162 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
163      165      170      175
165 GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG GGC CTC GCC      576
166 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
167      180      185      190
169 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA      624
170 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
171      195      200      205
173 CTC TTT GCT CTC TAATAG      642
174 Leu Phe Ala Leu
175      210
178 (2) INFORMATION FOR SEQ ID NO: 4:
180      (i) SEQUENCE CHARACTERISTICS:
181          (A) LENGTH: 212 amino acids
182          (B) TYPE: amino acid
183          (D) TOPOLOGY: linear
185      (ii) MOLECULE TYPE: protein
187      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
189 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
190 1      5      10      15
192 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
193      20      25      30
195 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
196      35      40      45
198 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
199      50      55      60
201 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
202 65      70      75      80
204 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
205      85      90      95
207 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
208      100      105      110
210 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
211      115      120      125
213 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
214      130      135      140
216 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
217 145      150      155      160
219 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
220      165      170      175

```

RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/973,025

TIME: 16:06:22

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

222 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
 223 180 185 190
 225 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
 226 195 200 205
 228 Leu Phe Ala Leu
 229 210

232 (2) INFORMATION FOR SEQ ID NO: 5:

234 (i) SEQUENCE CHARACTERISTICS:

235 (A) LENGTH: 795 base pairs

236 (B) TYPE: nucleic acid

237 (C) STRANDEDNESS: single

238 (D) TOPOLOGY: linear

240 (ii) MOLECULE TYPE: cDNA

242 (iii) HYPOTHETICAL: NO

C--> 244 (iv) ANTI-SENSE: NO

247 (ix) FEATURE:

248 (A) NAME/KEY: CDS

249 (B) LOCATION: 1..792

251 (ix) FEATURE:

252 (A) NAME/KEY: mat_peptide

253 (B) LOCATION: 1..789

256 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

258	ATG	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTT	ACA	TGC	GGC	TTC	GCC	GAC	CTC	48
259	Met	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp	Leu	
260	1			5				10						15			
262	GTG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCC	CTA	GGG	GGC	GCT	GCC	AGG	96
263	Val	Gly	Tyr	Ile	Pro	Leu	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg	
264			20					25						30			
266	GCC	CTG	GCG	CAT	GGC	GTC	CGG	GTT	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	144
267	Ala	Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	
268			35					40						45			
270	ACA	GGG	AAT	TTG	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTC	TTG	GCT	TTG	192
271	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	
272			50					55						60			
274	CTG	TCC	TGT	CTG	ACC	GTT	CCA	GCT	TCC	GCT	TAT	GAA	GTG	CGC	AAC	GTG	240
275	Leu	Ser	Cys	Leu	Thr	Val	Pro	Ala	Ser	Ala	Tyr	Glu	Val	Arg	Asn	Val	
276	65					70					75				80		
278	TCC	GGG	ATG	TAC	CAT	GTC	ACG	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT	GTG	288
279	Ser	Gly	Met	Tyr	His	Val	Thr	Asn	Asp	Cys	Ser	Asn	Ser	Ser	Ile	Val	
280				85						90					95		
282	TAT	GAG	GCA	GCG	GAC	ATG	ATC	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	336
283	Tyr	Glu	Ala	Ala	Asp	Met	Ile	Met	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	
284			100						105					110			
286	GTT	CGG	GAG	AAC	TCT	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACC	CCC	ACG		384
287	Val	Arg	Glu	Asn	Asn	Ser	Ser	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	
288			115					120						125			
290	CTC	GCA	GCT	AGG	AAC	GCC	AGC	GTC	CCC	ACC	ACG	ACA	ATA	CGA	CGC	CAC	432
291	Leu	Ala	Ala	Arg	Asn	Ala	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	
292		130						135						140			

RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/973,025

TIME: 16:06:22

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

```

294 GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGT TCC GCT ATG TAC GTG      480
295 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
296 145      150      155      160
298 GGG GAC CTC TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC      528
299 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile
300      165      170      175
302 TCG CCT CGC CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT      576
303 Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr
304      180      185      190
306 CCC GGC CAC ATA ACG GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC      624
307 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
308      195      200      205
310 TGG TCG CCT ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA      672
311 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
312      210      215      220
314 CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG      720
315 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
316 225      230      235      240
318 GGT CTC GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT      768
319 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
320      245      250      255
322 GTG ATG CTA CTC TTT GCT CCC TAATAG      795
323 Val Met Leu Leu Phe Ala Pro
324      260
327 (2) INFORMATION FOR SEQ ID NO: 6:
329 (i) SEQUENCE CHARACTERISTICS:
330 (A) LENGTH: 263 amino acids
331 (B) TYPE: amino acid
332 (D) TOPOLOGY: linear
334 (ii) MOLECULE TYPE: protein
336 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
338 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
339 1      5      10      15
341 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
342      20      25      30
344 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
345      35      40      45
347 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
348      50      55      60
350 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
351 65      70      75      80
353 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
354      85      90      95
356 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
357      100      105      110
359 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr
360      115      120      125
362 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
363      130      135      140

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/973,025

DATE: 11/30/2001

TIME: 16:06:23

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:64 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:85 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:107 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:244 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:402 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:470 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:539 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:590 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9
L:647 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:756 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:894 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:914 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:934 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:954 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:974 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:994 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1014 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1086 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:21
L:1158 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1281 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1411 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1548 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1684 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1821 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1841 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:1861 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2116 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2308 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2501 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2679 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:2859 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3101 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3447 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:3663 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:49
L:4727 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4747 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4769 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4790 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4810 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4830 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4850 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4870 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4890 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4910 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4930 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4950 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]

VERIFICATION SUMMARY

DATE: 11/30/2001

PATENT APPLICATION: US/09/973,025

TIME: 16:06:23

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\11302001\I973025.raw

L:4970 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:4990 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5010 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5030 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5050 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]
L:5070 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]